Lectin from Sainfoin (*Onobrychis viciifolia* Scop.). Complete Amino Acid Sequence†

Rozenn N. Kouchalakos,‡ Owen J. Bates,‡ Ralph A. Bradshaw,‡ and Kenneth D. Hapner*

ABSTRACT: The complete amino acid sequence of a lectin from sainfoin (*Onobrychis viciifolia* Scop. var. Eski) has been determined by sequential Edman analyses of the intact protein and peptides derived from digests with trypsin and thermolysin. Peptides were purified by pH fractionation, by gel filtration, and by cation-exchange and reverse-phase high-performance liquid chromatography. Seven segments of continuous sequence, accounting for the entire protein, were aligned through sequence comparison with several homologous leguminous lectins to give the final structure. Sainfoin lectin monomer, a glycoprotein which contains a single polypeptide chain of 236 amino acid residues with a molecular weight of 26 509, has amino- and carboxyl-terminal residues of alanine and threonine, respectively. A single residue of cysteine, located at position 33, is the only sulfur-containing amino acid present. Asparagine-118 is the single oligosaccharide attachment site. At least two apparent allelomorphic forms of the protein, having valine or isoleucine at position 49 in equal amounts, were detected. The amino acid sequence of sainfoin lectin exhibits circular permutation relative to that of the homologous protein concanavalin A.

Sainfoin lectin, prepared from seeds, is a 52 000 molecular weight glycoprotein that binds D-mannose (D-glucose) and specifically agglutinates trypsinized cat erythrocytes (Hapner & Robbins, 1979). The protein contains two identical, noncovalently associated polypeptide chains of molecular weight near 26 000. Each protein subunit contains one hexasaccharide unit attached via an N-glycosidic linkage to an asparagine residue (Namen & Hapner, 1979). Additional molecular characterization, to be described elsewhere, shows the protein to be a typical member of the D-mannose (D-glucose) binding family of leguminous lectins (Goldstein & Hayes, 1978). Other leguminous lectins with known primary structures are concanavalin A (Con A)[1] (Wang et al., 1975; Cunningham et al., 1975), favin (Hopp et al., 1982; Cunningham et al., 1979), and lentil lectin (Foriers et al., 1981). An incomplete structure for soybean agglutinin (Hemperly et al., 1982) and several N-terminal sequences (Foriers et al., 1979) have also been reported. This research, in which the complete amino acid sequence of sainfoin lectin was determined, was performed as part of an interdisciplinary study directed to improve sainfoin as a forage legume (Ditterline & Cooper, 1975) and to allow structural comparison with related proteins.

Materials and Methods

Protein Preparation. Sainfoin lectin was isolated from ground sainfoin seeds (gift from R. L. Ditterline, Montana State University) with a slight modification of the previous procedure (Hapner & Robbins 1979). The initial extraction step was made 0.1 M in D-glucose, and the ammonium sulfate fractionation step was omitted. Insoluble debris was removed by filtration and centrifugation, and the lectin was absorbed to the affinity matrix in batch fashion. D-Glucose was removed by dialysis prior to affinity absorption. Lectin concentration was determined by using $E_{280nm}^{1\%,1cm}$ = 14.3. The lyophilized protein was dissolved in 30% (v/v) formic acid prior to introduction into the cup of the sequencer.

S-Carboxymethylation and Succinylation. The single thiol group (Hapner & Robbins, 1979) of sainfoin lectin was alkylated with iodo[1-^{14}C]acetic acid according to Thomas et al. (1981). Succinylation was performed immediately after completion of the alkylation reaction. Solid succinic anhydride, in a 50-fold molar excess over the total protein lysine content, was added to the reaction solution at 22 °C in five 10-min intervals. The pH was held constant by the addition of 0.1 M NaOH. Reagents were removed by dialysis in H_2O, and the extent of lysine modification was determined by TNBS analysis (Habeeb, 1966).

Enzymatic Digests. Cleavage at arginine residues was accomplished by incubation of *S*-carboxymethylated and succinylated protein (2 mg/mL in 1.0 mM Tris-HCl, pH 8.0, 50 °C) with 1% (w/w) TPCK-treated trypsin. The pH was maintained constant by addition of 0.1 M NaOH. After base uptake had stopped, the reaction mixture was adjusted to and incubated at pH 2, 50 °C, for 1 h. The pH was then readjusted to 8.0, and the digestion was performed a second time with fresh TPCK-treated trypsin. The extent of arginine cleavage was estimated by analysis on an amino acid analyzer of the free arginine released after treatment of the peptides with DFP-treated carboxypeptidase B (Ambler, 1967). Digestion with thermolysin was performed on succinylated and S-carboxymethylated protein (5 mg/mL in 0.12 M ammonium bicarbonate–1 mM $CaCl_2$, pH 8.0, 22 °C) with 1% (w/w) thermolysin. After 1 h, the reaction mixture was applied to a column of Sephadex G-25. Carboxypeptidase Y digestion of isolated peptides was carried out in 0.1 M pyridine acetate buffer, pH 5.5 at 22 °C. Aliquots were removed at timed intervals, mixed with 0.2 M citrate buffer, pH 2.2, and analyzed directly on an amino acid analyzer. The method is generally that of Hayashi (1977).

† From the Department of Chemistry and Montana Agricultural Experiment Station, Montana State University, Bozeman, Montana 59717 (R.N.K. and K.D.H.), and the Department of Biological Chemistry, School of Medicine, Washington University, St. Louis, Missouri 63110 (O.J.B. and R.A.B.). *Received August 22, 1983.* This work was supported in part by grants from the U.S. Department of Agriculture (616-15-76) and the U.S. Public Health Service (AM 13362) and is published as Paper No. J 1449 from the Montana Agricultural Experiment Station. A preliminary communication has appeared (Kouchalakos & Hapner, 1982).

‡ Present address: Department of Biological Chemistry, California College of Medicine, University of California, Irvine, CA 92717.

[1] Abbreviations: TPCK, *N*-tosyl-L-phenylalanine chloromethyl ketone; DFP, diisopropyl fluorophosphate; HPLC, high-performance liquid chromatography; Pth, phenylthiohydantoin; Con A, concanavalin A; TNBS, trinitrobenzenesulfonic acid; Tris, tris(hydroxymethyl)aminomethane.

0006-2960/84/0423-1824$01.50/0 © 1984 American Chemical Society

Peptide Purification. The tryptic peptides were initially separated into soluble and insoluble fractions by centrifugation at 3000*g* after precipitation at pH 3.5 and 4 °C.

Peptides insoluble at pH 3.5 were redissolved in 0.01 M ammonium bicarbonate at pH 7.0 and subjected to gel filtration on Sephadex G-50. Peptides were detected by monitoring the absorbance at 230 nm. When necessary, further purification was achieved by reverse-phase HPLC on a 4.6 mm × 25 cm Beckman Ultrasphere ODS column. The mobile phase was 0.01 M ammonium bicarbonate, pH 7.0, and the mobile phase modifier was acetonitrile. The concentration of acetonitrile was rapidly increased to 20% (v/v) and then increased linearly from 20% to 35% in 30 min at a flow rate of 0.5 mL/min. Elution profiles were monitored at 280 nm.

Peptides soluble at pH 3.5 from a nonalkylated sample of lectin were lyophilized, redissolved in 30% (v/v) acetic acid, and subjected to cation-exchange chromatography on Dowex 50-X8 (AA15, Spinco Division, Beckman Instruments). The column was developed at 30 mL/h with pyridine acetate buffers (0.2 N, pH 3.1, to 2.0 N, pH 5.0) (Bradshaw et al., 1980) and monitored by the effluent reaction with ninhydrin following alkaline hydrolysis on an automated peptide analyzer as previously described (Hill & Delaney, 1967). In a separate sample, the peptide containing the radioactive *S*-(carboxymethyl)cysteine residue was initially purified by gel filtration using Sepadex G-25 with 1 mM acetic acid as eluant. Fractions containing radioactivity were concentrated by evaporation, and the peptide was further purified on a column of Sephadex LH20 using water as the eluant. The absorbance of the column effluent was monitored at 230 nm, and radioactivity was detected by liquid scintillation counting of aliquots from column fractions.

Thermolytic peptides were initially separated on a column of Sephadex G-25 in 0.02 M ammonium bicarbonate, pH 6.4, at a flow rate of 30 mL/h. The column effluent was monitored for absorbance at 230 nm. Aliquots of the emerging peaks were hydrolyzed and analyzed for arginine on an amino acid analyzer. The peaks that contained arginine were lyophilized and further purified on Dowex 50-X8 by using pyridine acetate buffers as described above. Peptides that contained arginine and whose amino acid composition suggested overlapping segments of previously sequenced tryptic peptides were lyophilized and submitted to sequence analysis.

Amino Acid Analysis. Samples were hydrolyzed in 6 N HCl under reduced pressure for 18–24 h in sealed glass tubes at 110 °C. Hydrolysates were dried in a desiccator or by rotary evaporation and analyzed on a Beckman 120C or Durrum D500 analyzer. No corrections were made for partial hydrolytic destruction of serine and threonine.

Carbohydrate Analysis. The carbohydrate-containing peptide was identified colorimetrically by the phenol–sulfuric acid method (Dubois et al., 1956) as modified by Misaki & Goldstein (1977).

Sequence Analysis. All polypeptides were sequenced automatically in a Beckman 890C sequencer using a 0.33 M Quadrol program (Thomas et al., 1981). Polybrene (Abbott) was added (~3 mg) to each sample prior to analysis. Phenylthiohydantions, after conversion from the phenylthiazolinones, were identified by a combination of thin-layer and gas chromatography or by reverse-phase HPLC (Thomas et al., 1981; Gordon et al., 1982). The phenylthiohydantoin of [^{14}C]-*S*-(carboxymethyl)cysteine was identified by liquid scintillation counting and by thin-layer chromatography.

Results

N-Terminal Sequence Analysis of Whole Protein. The entire sequence of sainfoin lectin is included in Figure 1. Edman degradation of the intact protein gave unambiguous identification of residues 1–26 and partial sequence information, of residues 27–39. Alanine was identified as the amino-terminal residue[2] and sequence positions 27–39 were subsequently confirmed through analysis of thermolytic and tryptic peptides.

Products of Trypsin Cleavage at Arginine. Sainfoin lectin which had been S-carboxymethylated and succinylated was extensively digested by trypsin. Treatment of the tryptic peptides with carboxypeptidase B typically released about 9.7 mol of arginine (of a theoretical maximum of 11) per mol of modified protein. Sainfoin lectin contains 12 arginine residues; however, one is present in the trypsin-resistant Arg-Pro sequence at positions 43–44. The extent of trypsin cleavage at susceptible arginine was therefore approximately 88%.

The modified protein and the tryptic peptides derived from it were soluble at pH 8.0. The peptides produced by tryptic cleavage at arginine are shown in Table I. At pH 3.5, the peptide digest precipitated, and 65% of the absorbancy at 280 nm was associated with the insoluble fraction. The soluble and insoluble fractions were separated by centrifugation and subjected to further purification.

The fraction insoluble at pH 3.5 was redissolved by adjusting the pH to 7.0. It was then applied to a column of Sephadex G50 from which three large peptides were isolated as shown in Figure 2A. Peptide T6 was the largest tryptic peptide produced, and it was isolated in pure form from the gel filtration column. It contained 46 amino acid residues and represented sequence positions 87–131 (Figure 1). Peptide T6 was a glycopeptide and accounted for 95% of the carbohydrate content of sainfoin lectin. The compositional and sequential analyses of T6 are incorporated into Table I. The carbohydrate attachment site was assigned to Asn-118 (Figure 1) because no Pth-amino acid was identified at that position and Asn-118 occurred in the unique sequence Ser-Asn-Glu-Thr which was previously identified as the site of glycosidic linkage (Namen & Hapner, 1979). Peptides T7 and T4 were repurified by reverse-phase HPLC (panels B and C, respectively, of Figure 2). They respectively contained 40 and 30 amino acids and accounted for residues 133–172 and 52–81 (Figure 1 and Table I). The three largest arginine peptides (T4, T6, T7) accounted for 116 (49%) of the amino acid residues in the protein. Three of the five tryptophan residues in sainfoin lectin were contained in T4 and T7 (Table I) and thus accounted for the 280-nm absorbance associated with the insoluble peptide fraction.

The pH 3.5 supernatant fraction that contained the soluble tryptic arginine peptides was chromatographed on Dowex 50-X8 as shown in Figure 3. The carboxyl-terminal peptide T12 lacked arginine and was not retained by the column. Six additional peptides, T3, T5, T8, T9, T10, and T11, were isolated in sufficiently pure form for sequence analysis. The peptides soluble at pH 3.5 ranged in size from 5 to 16 amino acid residues. Their sequence and composition are included in Figure 1 and Table I. Tryptic peptide T3 contained two Arg residues as a result of the resistant Arg-Pro sequence at positions 43–44, and it also gave equal amounts of Pth-Val and Pth-Ile at position 49. Peptide T2 was identified as an impure fraction (between T9 and T11, Figure 3) from the presence of half-cystine (composition given in Table I). In a separate sample, in which the cysteine had been converted

[2] Contrary to an earlier report (Hapner & Robbins, 1979), the amino terminus is not blocked.

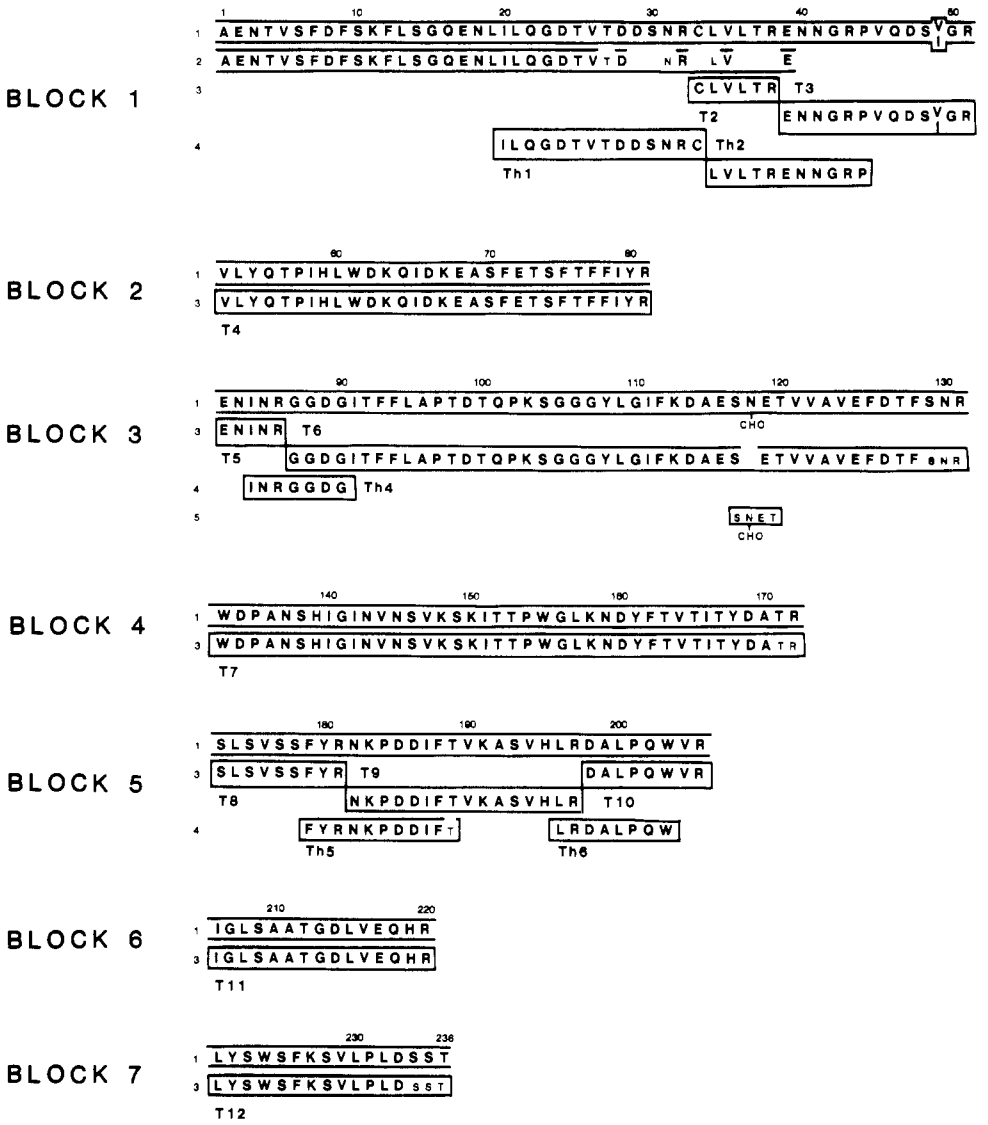

FIGURE 1: Summary proof of the primary structure of seven blocks of continuous sequence from sainfoin lectin. Each enclosed bar designates an isolated peptide. Large letters indicate positive Edman indentification, and small letters indicate tentative Edman identification or identification by carboxypeptidase analysis. The solid overlines indicate positive residue identification. Gaps in the overlines indicate tentative or no identification of that residue. The numbered lines (left) indicate the following: (1) final sequence; (2) amino-terminal sequence of intact protein; (3) sequences of tryptic (T) peptides; (4) sequences of thermolytic (Th) peptides; (5) sequence of Pronase glycopeptide (Namen & Hapner, 1979).

to [^{14}C]-*S*-(carboxymethyl)cysteine, peptide T2 was isolated from the other soluble tryptic peptides by gel filtration as shown in Figure 4. The peptide, which contains six amino acid residues, was completely sequenced from this sample, but no composition data were obtained for technical reasons.

A total of 11 major tryptic peptides were isolated (Table I). These peptides and the peptide represented by the amino-terminal 32 residues accounted for the complete amino acid composition of sainfoin lectin as indicated in Table II. Eight of the peptides (T2, T3, T4, T5, T8, T9, T10, and T11) were sequenced through their carboxyl-terminal arginine residues (Figure 1). Peptides T6, T7, and T12, respectively, gave positive amino acid identifications in 45 of 46, 38 of 40, and 13 of 16 sequential degradative cycles. Their carboxyl-terminal residues were assigned by carboxypeptidase analyses as indicated in Figure 1.

Thermolytic Peptides. Digestion of S-carboxymethylated and succinylated sainfoin lectin with thermolysin produced soluble peptides which were separated into nine fractions by gel filtration on a column of Sephadex G-25 (data not shown). Fractions that contained arginine were further purified by ion-exchange chromatography on Dowex 50-X8 in a pyridine acetate buffer system. Seven arginine-containing thermolytic peptides were isolated in pure form, and they were sequenced through their carboxyl termini. Five of these peptides, i.e., Th1, Th2, Th4, Th5, and Th6, provided useful sequence information and they are included in Figure 1. Two of the purified thermolytic peptides were derived from the carboxyl regions of tryptic peptides T3 and T11 and consequently contained no overlapping regions of sequence.

Alignment of Tryptic Peptides. N-Terminal sequence analyses of intact protein and tryptic and thermolytic peptides resulted in seven segments of "blocks" of continuous sequence as shown in Figure 1. Block 1 is composed of the amino-terminal region of the protein followed by tryptic peptides T2 and T3 whose positions were assigned by overlapping the thermolytic peptides Th1 and Th2. Although the amino-terminal region and T2 overlapped by a single (cysteine) amino acid residue, possible ambiguity is eliminated because no other cysteine residue is contained in the molecule. Block 2 is represented by peptide T4. Block 3 contains peptides T5 and T6 which were positioned by overlapping peptide Th4. Block 4 consists of peptide T7. Block 5 contains peptides T8, T9, and T10 whose positions were assigned by overlapping peptides

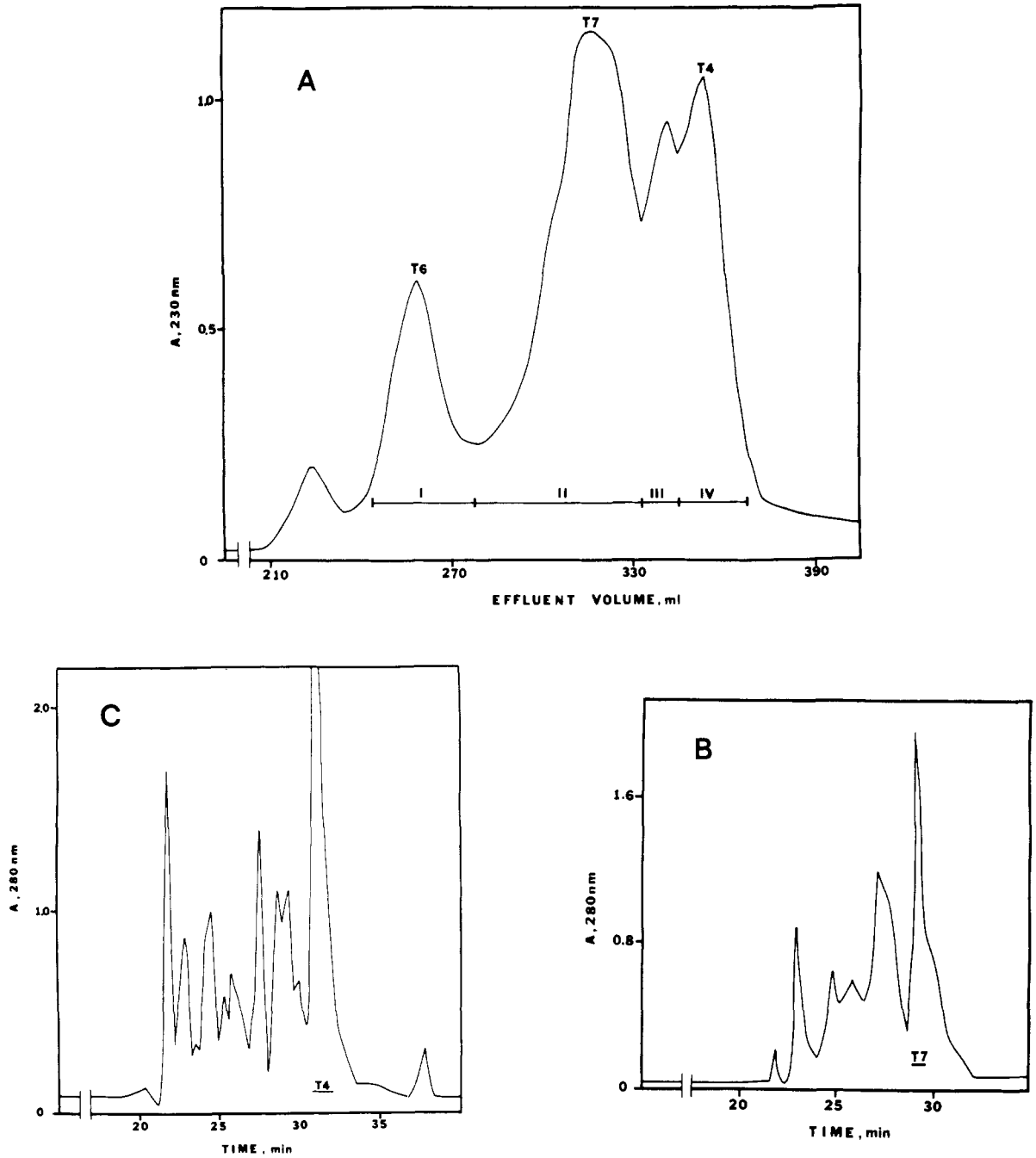

FIGURE 2: Purification of tryptic peptides insoluble at pH 3.5. (A) Elution profile of redissolved peptides (1.0 μmol) applied to a Sephadex G50 (2.5 × 120 cm) column. The column was developed at 30 mL/h, and 2-mL fractions were pooled as indicated. (B, C) Repurification of peptides T7 and T4 by reverse-phase HPLC. (For further details, see text.)

Th5 and Th6. Blocks 6 and 7 are represented by peptides T11 and T12, respectively.

Discussion

Assignment of Complete Amino Acid Sequence. Edman degradation analyses of intact protein and of protein-derived tryptic and thermolytic peptides gave seven blocks of continuous sequence as shown in Figure 1. Since no thermolytic peptides were isolated that contained overlapping sequences between individual blocks, it was necessary to align the total structure by other means. Blocks 1 and 7 were assigned, respectively, to the amino and carboxyl regions of the protein on the basis of results of whole protein sequence analysis and the lack of arginine in peptide T12. The five remaining blocks of sequence were aligned by structural homology with several related leguminous lectins as shown in Figure 5. When compared with the completed structures of favin (Hopp et al., 1982; Cunningham et al., 1979), lentil lectin (Foriers et al., 1981), and Con A (Wang et al., 1975; Cunningham et al., 1975) and the partial structure of soybean lectin (Hemperly et al., 1983), the five remaining blocks of sainfoin lectin sequence may be unambiguously aligned to give the complete sequence. The comparative homologous sequences in Figure 5 are arranged to maximize sequence identities. The amino-terminal regions of sainfoin lectin and favin, represented by block 1 in Figure 5 show 34% homology, expressed as amino acid identites/positions compared. Blocks 2, 3, 4, 5, 6, and 7 of the two structures show, respectively, 43, 52, 45, 32, 47, and 38% sequence homology. Similar degrees of homology are seen with the other lectins included in Figure 5. Overall homology between sainfoin lectin and the other sequenced proteins, expressed as percentage identities/positions compared,

Table I: Amino Acid Compositions[a] of Peptides Obtained by Arginine Cleavage of Sainfoin Lectin

	T2 33-38[b]	T3 39-51[b]	T4 52-81[b]	T5 82-86[b]	T6 87-132[b]	T7 133-171[b]	T8 172-180[b]	T9 181-196[b]	T10 197-204[b]	T11 205-219[b]	T12 220-234[b]
aspartic acid	0.7	2.7 (3)	2.0 (2)	1.7 (2)	6.0 (6)	7.0 (7)		3.0 (3)	1.2 (1)	1.1 (1)	1.0 (1)
threonine	1.1 (1)		2.6 (3)		4.5 (5)	5.0 (6)		0.9 (1)		1.0 (1)	0.8 (1)
serine	0.7	1.0 (1)	1.7 (2)		3.5 (3)	4.2 (3)	3.4 (4)	1.0 (1)		1.0 (1)	3.2 (5)
glutamic acid	0.7	1.9 (2)	4.0 (4)	0.8 (1)	4.0 (4)	1.0			1.3 (1)	2.0 (2)	
proline		1.0 (1)	0.9 (1)		ND[c] (2)	ND (2)		1.0 (1)	1.0 (1)		0.6 (1)
glycine	0.6	1.9 (2)			6.4 (7)	2.0 (2)		0.6	0.3	1.9 (2)	0.5
alanine	0.5		1.1 (1)		3.0 (3)	2.8 (2)		0.8 (1)	1.1 (1)		2.0 (2)
valine	0.9 (1)	1.5 (1.5)	1.1 (1)		2.6 (3)	4.0 (3)	1.0 (1)	1.5 (2)	0.8 (1)	0.9 (1)	0.6 (1)
half-cystine	0.2 (1)										
methionine											
isoleucine	0.4	0.5 (0.5)	2.9 (3)	1.1 (1)	1.9 (2)	3.8 (4)		0.9 (1)		1.1 (1)	
leucine	2.4 (2)	0.3	1.9 (2)		2.2 (2)	1.4 (1)	1.3 (1)	1.7 (1)	1.5 (1)	2.4 (2)	2.3 (3)
tyrosine			1.5 (2)		0.8 (1)	2.1 (2)	0.9 (1)				0.6 (1)
phenylalanine			3.5 (4)		4.6 (5)	2.0 (1)	0.9 (1)	0.8 (1)			0.7 (1)
tryptophan			ND (1)			ND (2)			ND (1)		ND (1)
lysine	0.6		1.9 (2)		1.9 (2)	3.0 (3)		1.0 (2)			0.5 (1)
histidine			0.7 (1)			1.5 (1)		0.6 (1)		0.9 (1)	
arginine	1.0 (1)	2.0 (2)	1.0 (1)	1.0 (1)	1.0 (1)	1.0 (1)	1.0 (1)	1.0 (1)	1.0 (1)	1.0 (1)	
residues no.	6	13	30	5	46	40	9	16	8	15	16
yield (%)	46	71	33	48	50	23	65	25	58	64	63

[a] Residues per molecule by amino acid analysis or (in parentheses) from the sequence (Figure 1). [b] Residue number. [c] ND, not determined.

Table II: Amino Acid Composition of Sainfoin Lectin

amino acid	whole protein hydrolysis[a]	sequence
lysine	11.4	11
histidine	3.6	4
arginine	11.4	12
aspartic acid	32.9	19
asparagine		14
threonine	19.8	21
serine	25.1	24
glutamic acid	18.5	10
glutamine		8
proline	9.4	9
glycine	16.1	15
alanine	12.0	11
half-cystine	0.9	1
valine	17.3	17.5
methionine	0	0
isoleucine	12.8	13.5
leucine	18.5	18
tyrosine	6.6	7
phenylalanine	16.0	16
tryptophan	5.0	5
	237.3[b]	236[b]

[a] Recalculated from Hapner & Robbins (1979). [b] Total.

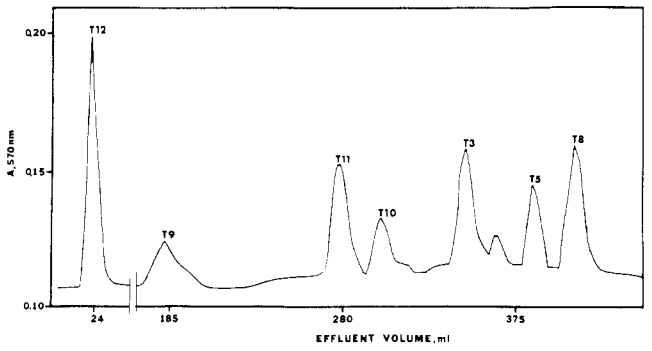

FIGURE 3: Elution profile of the separation of the tryptic peptides soluble at pH 3.5 on a column (0.9 × 20 cm) of Dowex 50-X8. Fractions of 3 mL were collected. Other details are described in the text.

is 42, 39, and 39% for favin, Con A, and lentil lectin, respectively (Table III).

Table III: Sequence Identities among Homologous Lectins[a]

	SL	Con A	favin	LL
SL		39%	42%	39%
Con A	95/246		38%	34%
favin	103/246	94/247		69%
LL	96/245	84/248	163/235	

[a] Percent identities (upper right quadrant) calculated from Figure 5 as identities/positions compared (lower left quadrant). Abbreviations: SL, sainfoin lectin; LL, lentil lectin; Con A, concanavalin A.

The complete sainfoin lectin structure contains 236 amino acid residues and has a calculated molecular weight of 26 509, both values in excellent agreement with previous measurements (Hapner & Robbins, 1979). Table II shows the comparative amino acid composition of sainfoin lectin as determined by amino acid and sequence analyses and further shows that the sequence, as determined here, accounts for all amino acids in the protein. The possibility does exist that a small tryptic peptide between blocks of sequence not covered by an overlap could have escaped detection. We feel this to be unlikely in light of the complete recovery of arginine from the protein (Table II).

The amino acid sequence determined here is in exact agreement with the amino-terminal sequence reported by Young et al. (1982). These workers studied a different variety of sainfoin and observed sequence heterogeneity at position 4 which was occupied by threonine or isoleucine. We observed threonine only in this position. We did, however, observe another instance of probable allelomorphic substitution at position 49, where valine and isoleucine were recovered in equal amounts. This substitution was confirmed by amino acid analysis of a chromatographically pure thermolytic peptide that contained residues 49–51. It is possible that other examples of sequence heterogeneity may have gone undetected, especially if present in low amounts. The fact that we did not observe sequence heterogeneity at position 4 is most likely explained by normal varietal genetic differences.

Sainfoin lectin is a member of a family of homologous leguminous lectins whose relatedness is indicated in Table III. The sainfoin lectin sequence shows approximately 40% identity

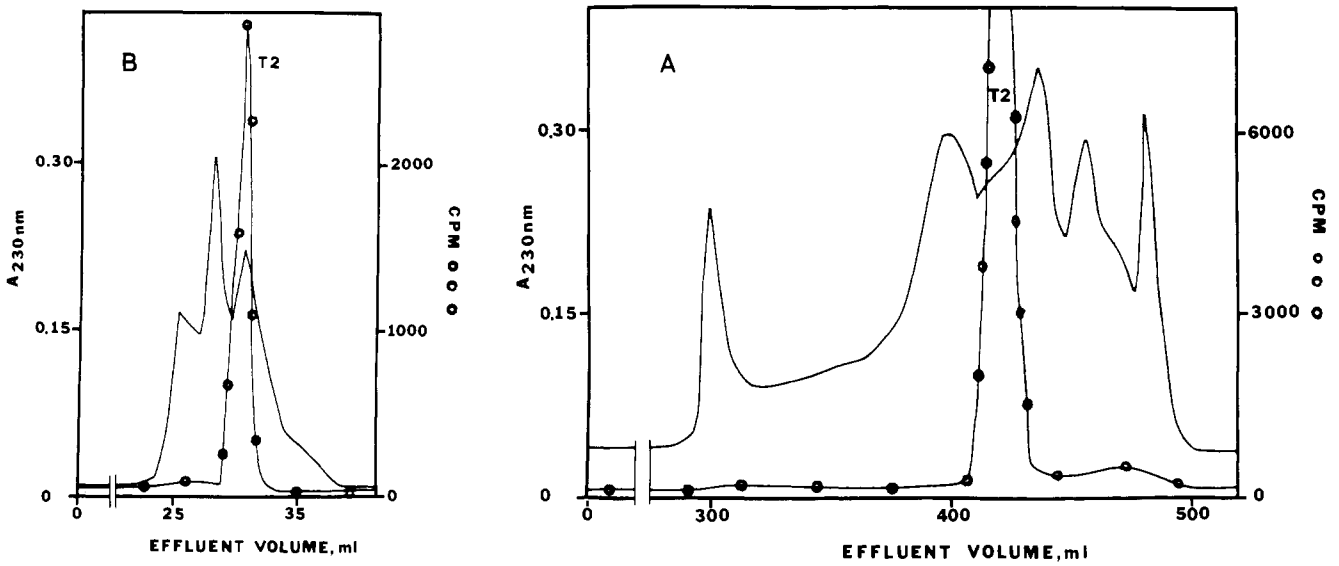

FIGURE 4: Purification of [^{14}C]-*S*-(carboxymethyl)cysteine peptide T2. (A) Elution profile of the gel filtration of tryptic peptides soluble at pH 3.5 (1.0 μmol) on a column (2.5 × 120 cm) of Sephadex G-25. The column was developed at 30 mL/h, and 2-mL fractions were collected. (B) Rechromatography of T2 on a column (0.9 × 95 cm) of Sephadex LH 20. Flow rate 30 mL/h, fraction size 2 mL.

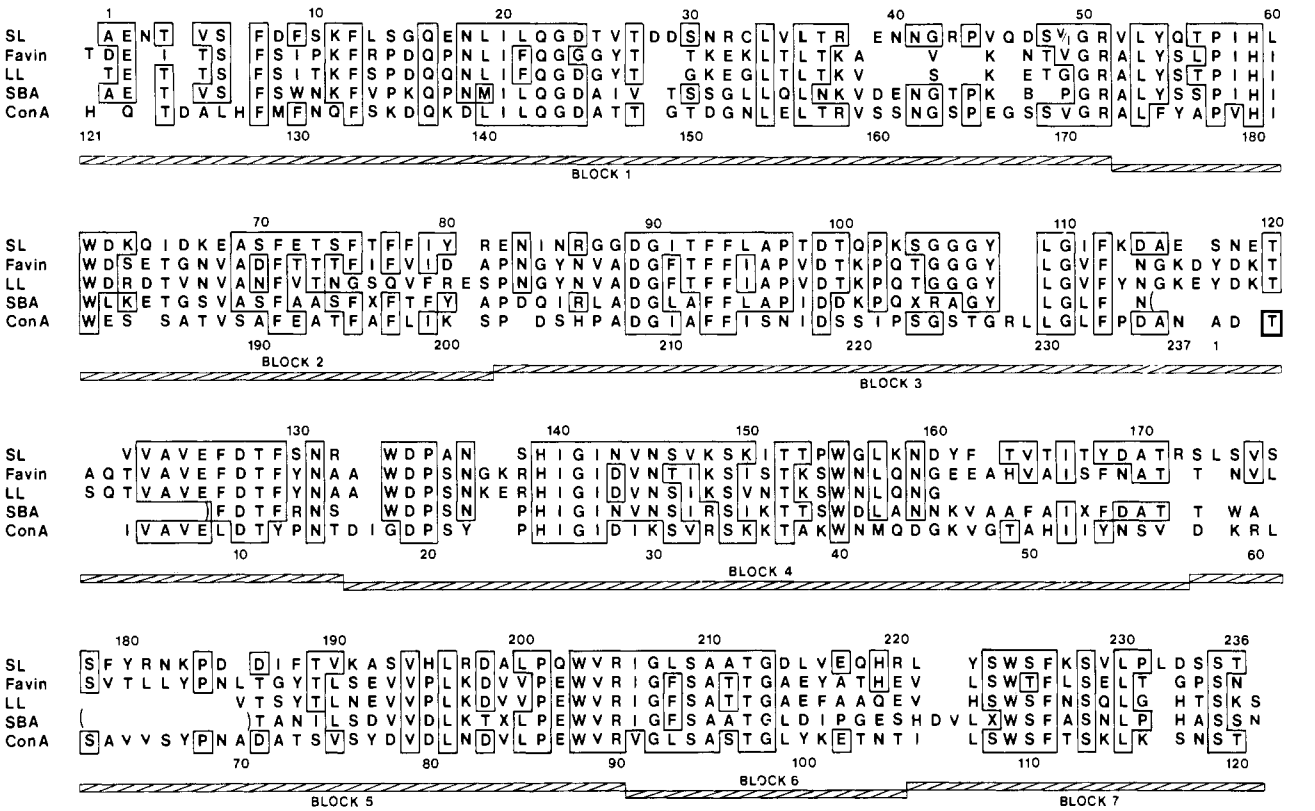

FIGURE 5: Final alignment of seven blocks of the sainfoin lectin sequence on the basis of homology. The hatched bars designate individual blocks of sequence (Figure 1). The boxes enclose amino acid residues identical with those in sainfoin lectin. Gaps are inserted in the sequence of the homologous proteins in order to maximize identities. Parentheses indicate incomplete sequence information in the SBA protein. The upper and lower numbering systems refer to the sequences of SL and Con A, respectively. Abbreviations: SL, sainfoin lectin; LL, lentil lectin; SBA, soybean agglutinin; Con A, concanavalin A.

with that of other completely sequenced lectins, i.e., Con A, favin, and lentil lectin. The greatest structural similarity among the proteins is seen with favin and lentil lectins which show 69% identity. This increased homology relative to the other sequenced lectins is consistent with the positioning of favin and lentil lectins within the same taxon whereas Con A and sainfoin lectin are more distantly related (Bell et al., 1978).

The carbohydrate attachment site was assigned as asparagine-118 on the basis of the following reasoning. The N-glycosidic nature of the sainfoin lectin glycosylation site was previously determined from a glycotetrapeptide isolated from a Pronase digest (Namen & Hapner, 1979). In the present work, the carbohydrate was identified in tryptic peptide T6 (Figure 1). Position 118 yielded no Pth-amino acid, as would be expected in the case of derivatized asparagine, and it was located in the unique sequence -Ser-X-Glu-Thr-, identical with the glycopeptide previously isolated. The position of asparagine-118, in sainfoin lectin, is homologous with the amino-terminal region of Con A (Figure 5) and correspondingly occupies an external position when compared with the three-

dimensional structure of Con A (Becker et al., 1975). The carbohydrate attachment position, among the homologous lectins which contain carbohydrate, is not conserved. Favin is glycosylated at Asn-168 (Hopp et al., 1982), and the soybean agglutinin carbohydrate site has been tentatively identified as Asn-75 (Becker et al., 1983), corresponding, respectively, to sainfoin lectin positions 169 and 76. Hence, the carbohydrate structure appears not to be involved in an important functional role, assuming all glycosylated lectins have similar roles within the plant.

Regions of the sainfoin lectin sequence corresponding to regions possessing certain three-dimensional structural characteristics in Con A (Becker et al., 1975; Reeke et al., 1975) are highly conserved (Hapner et al., 1983). Inspection of Figure 5 shows that amino acid side chains of Con A which bind Ca^{2+} and Mn^{2+}, i.e., glutamic acid-8, aspartic acids-10 and -19, asparagine-14, histidine-24, and the carbonyl oxygen of tyrosine-12, are conserved in all cases, except tyrosine-12 which is conservatively replaced by phenylalanine. The hydrophobic binding pocket and regions of β structure and monomer–monomer contact sites are well conserved. Amino acid residues involved in the carbohydrate binding site, in constrast, are poorly conserved, an observation consistent with the different saccharide binding affinities exhibited by these lectins (Goldstein & Hayes, 1978). The overall conservation of structurally significant regions of sequence suggests that the family of leguminous lectins will have similar tertiary structures, a conclusion also reached by computational methods (Olsen, 1983).

Among the five homologous lectins included in Figure 5, those of sainfoin, soybean, and Con A consist of a single polypeptide chain, whereas two polypeptide chains (α and β) compose the subunits of lentil lectin (Foriers et al., 1981) and favin (Hopp et al., 1982). Figure 5 shows that the amino-terminal regions of lectins from sainfoin and soybean are homologous with the amino-terminal portion of the β chains of favin and lentil lectin. Favin and lentil β chains correspond to sainfoin lectin residues 1–185 and 1–160, respectively, followed by the corresponding α chains that are similarly homologous with the carboxyl region of sainfoin lectin represented by residues 186–236. In order to maximize homology with Con A, it is necessary to align the carboxyl region of Con A, residues 121–273, with amino-terminal residues 1–116 of sainfoin lectin (Figure 5). Concanavalin A residues 1–120 consequently correspond to residues 117–236 of sainfoin lectin. This circular permutation of the Con A sequence relative to that of other leguminous lectins and its possible molecular basis have been described (Hemperly & Cunningham, 1983; Foriers et al., 1981).

The sainfoin lectin sequence reported here, along with the other homologous structures, supports the conclusion that the leguminous lectins are a family of related proteins whose structures have been evolutionarily conserved presumably for retention of important physiological functions.

Acknowledgments

We thank C. E. Bradshaw for technical assistance with aspects of this research.

References

Ambler, R. P. (1967) *Methods Enzymol. 11*, 155–166.

Becker, J. W., Reeke, G. N., Wang, J. L., Cunningham, B. A., & Edelman, G. M. (1975) *J. Biol. Chem. 250*, 1513–1524.

Becker, J. W., Hemperly, J. J., & Cunningham, B. A. (1983) in *Chemical Taxonomy, Molecular Biology, and Function of Plant Lectins* (Goldstein, I. J., & Etzlar, M. E., Eds.) pp 31–45, Alan R. Liss, New York.

Bell, E. A., Lackey, J. A., & Polhill, R. M. (1978) *Biochem. Syst. Ecol. 6*, 201–212.

Bradshaw, R. A., Bates, O. J., & Benson, J. R. (1980) *J. Chromatogr. 187*, 27–33.

Cunningham, B. A., Wang, J. L., Waxdal, M. J., & Edelman, G. M. (1975) *J. Biol. Chem. 250*, 1503–1512.

Cunningham, B. A., Hemperly, J. J., Hopp, T. P., & Edelman, G. M. (1979) *Proc. Natl. Acad. Sci. U.S.A. 76*, 3218–3222.

Ditterline, R. L., & Cooper, C. S. (1975) *Bull.—Mont., Agric. Exp. Stn. No. 681*, 1–23.

Dubois, M., Gillis, K. A., Hamilton, J. K., Rebers, P. A., & Smith, F. (1956) *Anal. Chem. 28*, 350–356.

Foriers, A., De Neve, R., & Strosberg, A. D., (1979) *Physiol. Veg. 17*, 597–606.

Foriers, A., Lebrun, E., Rapenbusch, R. V., de Neve, R., & Strasberg, A. D. (1981) *J. Biol. Chem. 256*, 5550–5560.

Goldstein, I. J., & Hayes, C. E. (1978) *Adv. Carbohydr. Chem. Biochem. 35*, 127–340.

Gordon, J. I., Smith, D. P., Andy, R., Alpers, D. H., Schonfeld, G., & Strauss, A. W. (1982) *J. Biol. Chem. 257*, 971–978.

Habeeb, A. F. S. A. (1966) *Anal. Biochem. 14*, 328–336.

Hapner, K. D., & Robbins, J. E. (1979) *Biochim. Biophys. Acta 580*, 186–197.

Hapner, K. D., Kouchalakos, R. N., & Bradshaw, R. A. (1983) in *Chemical Taxonomy, Molecular Biology, and Function of Plant Lectins* (Goldstein, I. J., & Etzlar, M. E., Eds.) pp 255–258, Alan R. Liss, New York.

Hayashi, R. (1977) *Methods Enzymol. 47*, 84–93.

Hemperly, J. J., & Cunningham, B. A. (1983) *Trends Biochem. Sci. (Pers. Ed.) 8*, 100–102.

Hemperly, J. J., Becker, J. W., & Cunningham, B. A. (1982) in *Proteins in Biology and Medicine* (Bradshaw, R. A., Hill, R. L., Tang, J., Liang, C., Tsao, T., Tsou, C., Eds.) pp 395–409, Academic Press, New York.

Hill, R. L., & Delaney, R. (1967) *Methods Enzymol. 11*, 339–351.

Hopp, T. P., Hemperly, J. J., & Cunningham, B. A. (1982) *J. Biol. Chem. 257*, 4473–4483.

Kouchalakos, R. N., & Hapner, K. D. (1982) *Fed. Proc., Fed. Am. Soc. Exp. Biol. 41*, 1180.

Misaki, A., & Goldstein, I. J. (1977) *J. Biol. Chem. 252*, 6995–6999.

Namen, A. E., & Hapner, K. D. (1979) *Biochim. Biophys. Acta 580*, 198–209.

Olsen, K. W. (1983) *Biochim. Biophys. Acta 743*, 212–218.

Reeke, G. N., Becker, J. W., & Edelman, G. M. (1975) *J. Biol. Chem. 250*, 1525–1547.

Thomas, K. A., Silverman, R. E., Jeng, I., Baglan, N. C., & Bradshaw, R. A. (1981) *J. Biol. Chem. 256*, 9147–9155.

Wang, J. L., Cunningham, B. A., Waxdel, M. J., & Edleman, G. M. (1975) *J. Biol. Chem. 250*, 1490–1502.

Young, N. M., Williams, R. E. Roy, C., & Yaguchi, M. (1982) *Can. J. Biochem. 60*, 933–941.